

**REMARKS**

In the final Office Action, the Examiner rejects claims 1, 5, 7, 9-13, 20-23, and 25-28 under 35 U.S.C. § 103(a) as allegedly unpatentable over U.S. Patent Application Publication No. 2004/0177015 to Galai et al. (hereinafter "GALAI") in view of U.S. Patent Application Publication No. 2004/0158429 to Bary et al. (hereinafter "BARY"); rejects claims 3, 8, and 15-19 under 35 U.S.C. § 103(a) as allegedly unpatentable over GALAI in view of BARY, and further in view of alleged Applicant's Admitted Prior Art (hereinafter "AAPA"); and rejects claims 4, 14, 24, and 29 under 35 U.S.C. § 103(a) as allegedly unpatentable over GALAI in view of BARY, and further in view of U.S. Patent No. 6,952,730 to Najork et al. (hereinafter "NAJORK). Applicant respectfully traverses these rejections.<sup>1</sup>

By way of this Amendment, Applicant proposes amending claims 1, 3-7, 9-13, 15-16, 18-22, and 25-29 to improve form. No new matter would be added by the amendment. Claims 1, 3-5, and 7-29 are pending.

Claims 1, 5, 7, 9, 10-13, 20-23, and 25-28 stand rejected under 35 U.S.C. § 103(a) as allegedly unpatentable over GALAI in view of BARY. Applicant respectfully traverses this rejection.

Independent claim 1, amended as proposed, is directed to a method that includes extracting a set of uniform resource locators (URLs) from one document or from multiple documents associated with a single web host, identifying sub-strings occurring in multiple URLs in the set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, generating a clean set of URLs from the set of

---

<sup>1</sup> As Applicant's remarks with respect to the Examiner's rejections are sufficient to overcome these rejections, Applicant's silence as to assertions by the Examiner in the Office Action or certain requirements that may be applicable to such rejections (e.g., whether a reference constitutes prior art, reasons to modify a reference and/or to combine references, assertions as to dependent claims, etc.) is not a concession by Applicant that such

URLs by removing the session identifiers, and determining when at least one particular URL has already been crawled based on a comparison of the particular URL to the clean set of URLs.

GALAI and BARY, whether taken alone or in any reasonable combination, do not disclose or suggest this combination of features.

For example, GALAI and BARY do not disclose or suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1. The Examiner relies on paragraphs [0005], [0013], [0023], [0067] of GALAI and on paragraphs [0184], [0196], and [0205] of BARY for allegedly disclosing "locating session identifiers in the set of URLs extracted as sub-strings that occur in multiple URLs of a web site" (final Office Action, p. 3). Applicant submits that the above sections of GALAI and BARY do not disclose or suggest the above feature of amended claim 1.

Paragraph [0005] of GALAI discloses:

However, many Web pages today are provided as dynamic Web pages, which are created in real time or "on the fly" from a plurality of components stored in a database. Dynamic Web pages are created upon submission of a query by a user, which determines the identity of the components to be retrieved and assembled into the Web page. For example, a URL for a dynamic Web page, if it exists, may appear as follows: <http://domain.com/search.asp?p1=v1&p2=v2>. The term "search.asp" is a name of an application which should be invoked, followed by a "?" sign, and a list of parameters and their values. Many autonomous software search programs are designed to ignore such links, since automatically following this type of link may cause an infinite recursion which the autonomous software program cannot properly handle. Thus, dynamic Web pages are often not indexed (by using filters to reject such Web pages automatically during the indexing process), or even "un-indexable" due to the fact that the only way to generate this page is by submitting a query through a form and not through a regular hyperlink used by search engines to locate new pages.

This section of GALAI discloses that many web pages are provided as dynamic web pages created in real time. An example is given, which includes the term "search.asp", followed by a "?" sign, and a list of parameters. This section of GALAI discloses that many software programs

are designed to ignore such links, and thus dynamic web pages are not indexed. This section of GALAI does not disclose or suggest session identifiers. Even if it is deemed reasonable that the parameters disclosed by this section of GALAI can be interpreted as session identifiers, a point Applicant does not concede, this section of GALAI does not disclose identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers. In fact, his section of GALAI does not disclose or even remotely suggest sub-strings that occur in multiple URLs of a set of URLs. Therefore, this section of GALAI cannot disclose or suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1.

Paragraph [0013] of GALAI discloses:

The removal of such non-essential code is preferably adjusted to a particular structure of Web pages or other type of document. Such a structure may optionally be found on a single Web site or other entity served by a particular Web server and/or dynamic Web page construction process or template. Such adjustment is most preferably performed by initially learning the structure of the Web pages, optionally by automatically scanning a plurality of Web pages produced with the same structure and/or by the same construction process. Such automatic scanning may also optionally include a statistical analysis of the Web pages, in order to infer extraction rules for such non-essential code. These extracting rules are optionally and more preferably based on statistical models, which determine the probability and/or the likelihood of a specific element of the page to be considered essential. As previously described, these Web pages may optionally have the same template, for example. The present invention then preferably detects repeated patterns in the Web page, more preferably by parsing the HTML code.

This section of GALAI discloses removing non-essential code from a web page by learning the structure of a web page by scanning a plurality of web pages produced with the same structure or by the same construction process. The extracting rules for removing non-essential code may be based on statistical models, which determine the probability of a specific element of the page being essential. Repeated patterns on a page may be detected, by preferably parsing the HTML code. This section of GALAI does not disclose or suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers. In fact, this section of GALAI does not

disclose or even remotely suggest sub-strings that occur in multiple URLs of a set of URLs.

Instead, this section of GALAI discloses parsing HTML code of a web page to detect repeated patterns. Parsing HTML code of a web page to detect patterns would not result in identifying sub-strings occurring in multiple URLs of a set of URLs. Therefore, this section of GALAI cannot disclose or suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1.

Paragraph [0023] of GALAI discloses:

The above process is preferably executed once per URL structure, and the normalization instructions are then applied on each URL with the same structure. The term "URL structure" preferably includes the same parameters, repeated for each such structure. The redundant parameters are preferably removed automatically before the Web page is retrieved and indexed by the search engine.

This section of GALAI discloses executing the previously disclosed process (paragraphs [0019-0022] of GALAI) of retrieving a first web page with a URL, removing a parameter from the URL, retrieving a second web page with the reduced URL, and comparing the first and second web pages to determine if they are similar. If the first and second web pages are determined to be similar, the parameter which was removed from the URL is determined to be redundant and is removed from the URL before the URL is indexed. This section of GALAI does not disclose or suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers. In fact, this section of GALAI does not disclose or even remotely suggest sub-strings that occur in multiple URLs of a set of URLs. Therefore, this section of GALAI cannot disclose or suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1.

Paragraph [0067] of GALAI discloses:

As for the previous embodiment, more preferably, the operation of the present invention is adjusted to a particular structure of Web pages, as may optionally be found on a single Web site or other entity served by a particular Web server and/or dynamic Web page construction process or template. Such adjustment is most preferably performed by initially learning the structure of the Web pages, optionally by automatically scanning a plurality of Web pages produced with the same or similar structure. As previously described, these Web pages may optionally have the same originating template and/or may optionally be generated by the same construction process, for example. The present invention then learns how to detect and extracts specific elements, or fields, from the page, optionally assigning attributes to each field and optionally associating each field with an information object or an attribute of an information object defined in an information schema. The attributes of the fields are preferably defined either automatically or manually per set of pages that have the same or similar structure, and preferably are derived from the information schema. As previously described, these Web pages may optionally have the same originating template, for example.

This section of GALAI discloses scanning a plurality of web pages produced with the same or similar structure to extract specific elements, or fields from the page, and associating each field with an information object or an attribute of an information object. This section of GALAI does not disclose or suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1.

In the Response to Arguments section of the final Office Action, The Examiner alleges that “Galai’s system must have the capability to ‘identify’ session identifiers (parameter) in order to remove the session identifier in the URL” (final Office Action, p. 2). Contrary to the Examiner’s allegation, however, GALAI does not disclose or suggest removing session identifiers in a URL. The Examiner did not reference a specific section of GALAI that allegedly discloses removing session identifiers in URL. Instead, GALAI discloses normalizing a URL by removing redundant parameters, where a parameter is any divisible subunit of a URL (GALAI, paragraph [0019]). GALAI determines that a parameter is redundant by retrieving the same web page with and without the parameter, and comparing the two web pages to determine whether the content is the same. If the content is determined to be sufficiently similar for the two web pages, the parameter is determined to be redundant and removed before indexing the URL. The

Examiner has not shown that removing a redundant parameter is equivalent to removing a session identifier.

Nevertheless, even if it is deemed reasonable that GALAI discloses removing session identifiers in a URL, a point Applicant does not concede, this does not mean it is obvious that GALAI discloses a particular method of identifying session identifiers. In fact, the particular method used by GALAI to determine whether a parameter is redundant includes retrieving a first web page with a URL, removing a parameter from the URL, retrieving a second web page with the reduced URL, and comparing the first and second web pages to determine if they are similar. If the first and second web pages are determined to be similar, the parameter which was removed from the URL is determined to be redundant and is removed from the URL before the URL is indexed (paragraphs [0019-0022] of GALAI). GALAI does not disclose or suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1.

The Examiner further relies on paragraphs [0184], [0196], and [0205] of BARY for allegedly identifying session identifiers (final Office Action, p. 3). Applicant submits that the alleged method of identifying session identifiers disclosed by BARY is unrelated to the above feature of amended claim 1.

Paragraph [0184] of BARY discloses:

2. the presence of session identifiers (session id's). A session id is a variable name within the URL that changes the characters in the URL string, but has no impact on how the URL traverses the Internet to arrive at the desired location; and

This section of BARY discloses that a session identifier is a variable name within a URL that changes characters in the URL string but has no impact on how the URL traverses the Internet to

arrive at the desired location. This section of BARY does not disclose or even remotely suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1.

Paragraph [0196] of BARY discloses:

3. option indicates that URLs have the following to identify sessions: sid, Sessionid, refer, and delimiters "&" and "\_". (i.e. delete all characters after "SID").

This section of BARY discloses that URLs use "sid", "sessionid", "refer", and "&" and "\_" delimiters to identify sessions. This section of BARY does not disclose or even remotely suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1.

Paragraph [0205] of BARY discloses:

Currently the session id is searched for within the entire URL so if the session id variable happens to be in the path then the URL will be stripped early. If the Web administrator had an option to identify a character that identified the beginning of any session variables then they could define where search started. In most sites this would be defaulted to the "?" character. To implement this the session id could be searched in the URL from anything following this character.

This section of BARY discloses searching for a session ID and stripping it from the URL. This section of BARY also discloses searching for the "?" character to identify session identifiers.

This section of BARY does not disclose or even remotely suggest identifying sub-strings occurring in multiple URLs in a set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs, as recited in amended claim 1.

With regard to reasons for combining GALAI and BARY, the Examiner alleges (final Office Action, p. 4):

The act alone (removing the session identifier) shows that it must be obvious the system has capability to identify session identifier in order to remove the session identifier. Bary, demonstrate the obviousness by disclosing the steps of how to identify the session identifier.

Applicant disagrees with the Examiner's allegation. Neither GALAI nor BARY disclose the specific method of identifying session identifiers recited in amended claim 1. GALAI discloses identifying redundant parameters by comparing two versions of the same web page, and BARY discloses identifying session identifiers by looking for specific terms or characters. Neither GALAI nor BARY disclose or suggest identifying sub-strings occurring in a set of URLs as session identifiers, based on a predetermined rule and based on the sub-strings occurring in multiple URLs of a set of URLs, as recited in amended claim 1.

For at least the foregoing reasons, Applicant submits that claim 1 is patentable over GALAI and BARY, whether taken alone or in any reasonable combination. Accordingly, Applicant respectfully requests that the rejection of claim 1 under 35 U.S.C. § 103(a) based on GALAI and BARY be reconsidered and withdrawn.

Claims 5, 7, and 9 depend from claim 1. Therefore, these claims are patentable over GALAI and BARY, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 1. Accordingly, Applicant respectfully requests that the rejection of claims 5, 7, and 9 under 35 U.S.C. § 103(a) based on GALAI and BARY be reconsidered and withdrawn.

Independent claims 10, 20, and 25 recite features similar to, yet possibly of different scope than, the features recited above with respect to claim 1. Therefore, these claims are patentable over GALAI and BARY, whether taken alone or in any reasonable combination, for at least reasons similar to the reasons set forth above with respect to claim 1. Accordingly,



Applicant respectfully requests that the rejection of claims 10, 20, and 25 under 35 U.S.C. § 103(a) based on GALAI and BARY be reconsidered and withdrawn.

Claims 11-13 depend from claim 10. Therefore, these claims are patentable over GALAI and BARY, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 10. Accordingly, Applicant respectfully requests that the rejection of claims 11-13 under 35 U.S.C. § 103(a) based on GALAI and BARY be reconsidered and withdrawn.

Claims 21-23 depend from claim 20. Therefore, these claims are patentable over GALAI and BARY, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 20. Accordingly, Applicant respectfully requests that the rejection of claims 21-23 under 35 U.S.C. § 103(a) based on GALAI and BARY be reconsidered and withdrawn.

Claims 26-28 depend from claim 25. Therefore, these claims are patentable over GALAI and BARY, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 25. Accordingly, Applicant respectfully requests that the rejection of claims 26-28 under 35 U.S.C. § 103(a) based on GALAI and BARY be reconsidered and withdrawn.

Claims 3, 8, and 15-19 stand rejected under 35 U.S.C. § 103(a) as allegedly unpatentable over GALAI in view of BARY, and further in view of AAPA. Applicant respectfully traverses this rejection.

Claims 3 and 8 depend from claim 1. Without acquiescing in the Examiner's rejection and assuming that AAPA is in fact prior art (a point that Applicant does not concede), Applicant submits that AAPA does not overcome the deficiencies of GALAI and BARY set forth above

with respect to claim 1. Therefore, claims 3 and 8 are patentable over GALAI, BARY, and AAPA, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 1. Accordingly, Applicant respectfully requests that the rejection of claims 3 and 8 under 35 U.S.C. § 103(a) based on GALAI, BARY, and AAPA be reconsidered and withdrawn.

Independent claim 15 recites features similar to, yet possibly of different scope than, features recited above with respect to claim 1. Without acquiescing in the Examiner's rejection and assuming that AAPA is in fact prior art (a point that Applicant does not concede), Applicant submits that AAPA does not overcome the deficiencies of GALAI and BARY set forth above with respect to claim 1. Therefore, claim 15 is patentable over GALAI, BARY, and AAPA, whether taken alone or in any reasonable combination, for at least reasons similar to the reason set forth above with respect to claim 1. Accordingly, Applicant respectfully requests that the rejection of claim 15 under 35 U.S.C. § 103(a) based on GALAI, BARY, and AAPA be reconsidered and withdrawn.

Claims 16-19 depend from claim 15. Therefore, claims 16-19 are patentable over GALAI, BARY, and AAPA, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 15. Accordingly, Applicant respectfully requests that the rejection of claims 16-19 under 35 U.S.C. § 103(a) based on GALAI, BARY, and AAPA be reconsidered and withdrawn.

Claims 4, 14, 24, and 29 stand rejected under 35 U.S.C. § 103(a) as allegedly unpatentable over GALAI in view of BARY, and further in view of NAJORK. Applicant respectfully traverses this rejection.

Claim 4 depends from claim 1. Without acquiescing in the Examiner's rejection, Applicant submits that NAJORK does not overcome the deficiencies of GALAI and BARY set forth above with respect to claim 1. Therefore, claim 4 is patentable over GALAI, BARY, and NAJORK, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 1. Accordingly, Applicant respectfully requests that the rejection of claim 4 under 35 U.S.C. § 103(a) based on GALAI, BARY, and NAJORK be reconsidered and withdrawn.

Claim 14 depends from claim 10. Without acquiescing in the Examiner's rejection, Applicant submits that NAJORK does not overcome the deficiencies of GALAI and BARY set forth above with respect to claim 10. Therefore, claim 14 is patentable over GALAI, BARY, and NAJORK, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 10. Accordingly, Applicant respectfully requests that the rejection of claim 14 under 35 U.S.C. § 103(a) based on GALAI, BARY, and NAJORK be reconsidered and withdrawn.

Claim 24 depends from claim 20. Without acquiescing in the Examiner's rejection, Applicant submits that NAJORK does not overcome the deficiencies of GALAI and BARY set forth above with respect to claim 20. Therefore, claim 24 is patentable over GALAI, BARY, and NAJORK, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 20. Accordingly, Applicant respectfully requests that the rejection of claim 24 under 35 U.S.C. § 103(a) based on GALAI, BARY, and NAJORK be reconsidered and withdrawn.

Claim 29 depends from claim 25. Without acquiescing in the Examiner's rejection, Applicant submits that NAJORK does not overcome the deficiencies of GALAI and BARY set

forth above with respect to claim 25. Therefore, claim 29 is patentable over GALAI, BARY, and NAJORK, whether taken alone or in any reasonable combination, for at least the reasons set forth above with respect to claim 25. Accordingly, Applicant respectfully requests that the rejection of claim 29 under 35 U.S.C. § 103(a) based on GALAI, BARY, and NAJORK be reconsidered and withdrawn.

Applicant respectfully requests that this proposed amendment under 37 C.F.R. § 1.116 be entered, placing the application in condition for allowance. In addition, Applicant respectfully submits that entry of this proposed amendment would place the application in better form for appeal in the event that the application is not allowed. If the Examiner does not believe that the claims are in condition for allowance, the Examiner is urged to contact the undersigned agent to expedite prosecution of this application.

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY SNYDER, L.L.P.

By: /Viktor Simkovic, Reg. No. 56012/  
Viktor Simkovic  
Registration No. 56,012

Date: July 11, 2008

11350 Random Hills Road  
Suite 600  
Fairfax, Virginia 22030  
(571) 432-0800 main  
(571) 432-0899 direct  
Customer Number: 44989